

「知性への誘い」(Cクラス)

講義ノート

谷崎 久志
大阪大学・経済学部

2018/06/22-29

目次

	5	ダミー変数	9
	5.1	異常値ダミー	9
	5.2	構造変化ダミー	10
	5.3	季節ダミー	10
	5.4	地域差ダミー	10
	5.5	男女別ダミー	10
	6	関数型について	10
	7	需要関数の計算と解釈(レポート, 締め切り 7月6日 PM17:00 まで厳守)	13
	7.1	データの入手方法	13
	7.2	例: 米の需要関数	14
1	1	計量経済学について	1
1.1	1	例1: 国の消費関数	1
1.2	1	例2: 日本酒の需要関数	1
2	2	回帰分析	2
2.1	2	重要な公式	2
2.2	2	データについて	2
3	2	最小二乗法について: 単回帰モデル	2
3.1	2	最小二乗法と回帰直線	2
3.2	2	切片 α と傾き β の求め方	2
3.3	4	残差 \hat{u}_i の性質について	4
3.4	4	決定係数 R^2 について	4
3.5	6	決定係数の比較	6
3.6	7	まとめ	7
4	7	最小二乗法について: 重回帰モデル	7
4.1	8	決定係数 R^2 と自由度修正済み決定係数 \bar{R}^2 について	8

● この講義ノートは,
<http://www2.econ.osaka-u.ac.jp/~tanizaki/class/2018>
からダウンロード可。

〔講義題目 (テーマ)〕

経済学における実証分析の方法

〔講義概要〕

経済学の基本は需要・供給ですが、こうした経済理論に対して実際の経済活動のデータを当てはめ、理論と現実が整合的かどうかを統計的に確かめるのが実証分析と呼ばれるものです。この一連の分析手順と用いる手法を、具体的にデータを用いながらお話しします。

1 計量経済学について

- 経済理論 (ミクロ経済, マクロ経済, 財政, 金融, 国際経済, ...)
- データ (GNP, 消費, 投資, 金利, 為替レート, ...)

計量経済学 \Rightarrow 経済理論が現実になり立つものかどうかを、データを用いて、統計的に検証する。

1.1 例 1: 国の消費関数

$$C = f(Y)$$

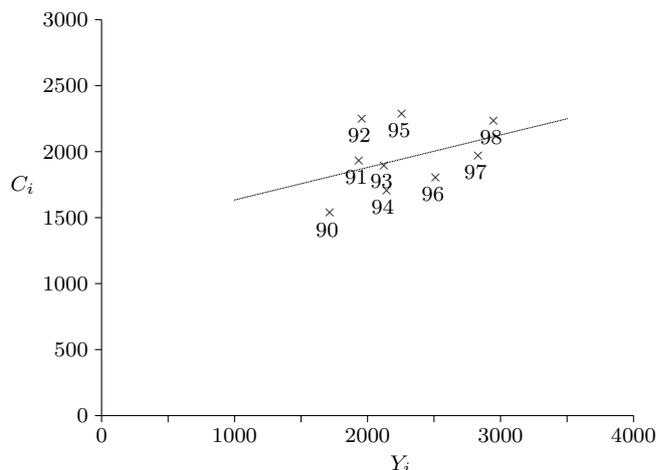
ただし, C は消費, Y は所得。

1. $Y \nearrow \Rightarrow C \nearrow$
2. $\frac{dC}{dY}$ = 限界消費性向 = 所得 1 円増加で消費が何円増加するか
3. すなわち, $\frac{dC}{dY} > 0$

モデルの定式化

1. $C = a + bY$
2. $b = \frac{dC}{dY}$ = 限界消費性向
3. a = 基礎消費 ($Y = 0$ のときに必要な消費)
4. 符号条件: $a > 0, b > 0$ (しかも, $1 > b$)

図 1: 消費 (C_i) と所得 (Y_i)



1. $\times \rightarrow$ 実際のデータ
 2. $(Y_i, C_i) \Rightarrow t$ 期のデータ, i.e., $i = 1, 2, \dots, 9$
 3. $i = 1 \Rightarrow$ 1990 年,
 $i = 2 \Rightarrow$ 1991 年,
...,
 $i = 9 \Rightarrow$ 1998 年,
1. 実際のデータを用いて, a, b を求める。
 2. a, b を求める \equiv 現実の経済構造を求める
 3. その結果, もし $a > 0, 1 > b > 0$ なら, 経済理論は現実経済を説明していると言える。

1.2 例 2: 日本酒の需要関数

$$Q = f(Y, P_1, P_2)$$

ただし, Q は日本酒の需要量, Y は所得, P_1 は日本酒の価格, P_2 は洋酒の価格。

1. $Y \nearrow \Rightarrow Q \nearrow,$
 $P_1 \nearrow \Rightarrow Q \searrow,$
 $P_2 \nearrow \Rightarrow Q \nearrow$
2. $\frac{\partial Q}{\partial Y} > 0, \frac{\partial Q}{\partial P_1} < 0, \frac{\partial Q}{\partial P_2} > 0$
3. 日本酒と洋酒は代替財

4. モデルの定式化 (A)

$$Q = a + b_1 Y + b_2 P_1 + b_3 P_2$$

5. Q, Y, P_1, P_2 を用いて, a, b_1, b_2, b_3 を求める (日本酒の需要構造を求める)。

6. 符号条件: $b_1 > 0, b_2 < 0, b_3 > 0, a ?$

7. t 期のデータ (Q_i, Y_i, P_{1i}, P_{2i})

8. n 組のデータ, i.e., $i = 1, 2, \dots, n$

9. モデルの定式化 (B)

$$Q = a + b_1 Y + b_2 \frac{P_1}{P_2}$$

符号条件: $b_1 > 0, b_2 < 0$

10. モデルの定式化 (C)

$$\log(Q) = a + b_1 \log(Y) + b_2 \log\left(\frac{P_1}{P_2}\right)$$

符号条件: $b_1 > 0, b_2 < 0$

11. モデル (A), (B), (C) のどれが最も現実的かを得られた結果から判断する。

2 回帰分析

2.1 重要な公式

$$1. \sum_{i=1}^n X_i = n\bar{X}$$

$$2. \sum_{i=1}^n (X_i - \bar{X}) = 0$$

$$3. \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

$$4. \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y} = \sum_{i=1}^n (X_i - \bar{X})Y_i = \sum_{i=1}^n (Y_i - \bar{Y})X_i$$

2.2 データについて

1. タイム・シリーズ (時系列)・データ: 添え字 i が時間を表す (第 i 期)。 t を添え字に使う場合も多い。

2. クロス・セクション (横断面)・データ: 添え字 i が個人や企業を表す (第 i 番目の家計, 第 i 番目の企業)。

3 最小二乗法について: 単回帰モデル

最小二乗法とは, 線型モデルの係数の値をデータから求める時に用いられる手法である。

3.1 最小二乗法と回帰直線

$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ のように n 組のデータがあり, X_i と Y_i との間に以下の線型関係を想定する。

$$Y_i = \alpha + \beta X_i,$$

X_i は説明変数, Y_i は被説明変数, α, β はパラメータとそれぞれ呼ばれる。

上の式は回帰モデル (または, 回帰式) と呼ばれる。切片 α と傾き β をデータ $\{(X_i, Y_i), i = 1, 2, \dots, n\}$ から推定することを考える。

ある基準の下で, α と β の推定値が求められたとしよう。それぞれ, $\hat{\alpha}$ と $\hat{\beta}$ とする。データ $\{(X_i, Y_i), i = 1, 2, \dots, n\}$ と直線との関係は,

$$Y_i = \hat{\alpha} + \hat{\beta} X_i + \hat{u}_i,$$

となる。すなわち, 実際のデータ Y_i と直線上の値 $\hat{\alpha} + \hat{\beta} X_i$ との間には, 誤差 \hat{u}_i (残差と呼ばれる) が生じる。

3.2 切片 α と傾き β の求め方

α, β のある推定値を $\hat{\alpha}, \hat{\beta}$ としよう。次のような関数 $S(\hat{\alpha}, \hat{\beta})$ を定義する。

$$S(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

これは残差平方和と呼ばれる。

このとき,

$$\min_{\hat{\alpha}, \hat{\beta}} S(\hat{\alpha}, \hat{\beta})$$

となるような $\hat{\alpha}, \hat{\beta}$ を求める (最小自乗法)。最小化のためには、

$$\frac{\partial S(\hat{\alpha}, \hat{\beta})}{\partial \hat{\alpha}} = 0, \quad \frac{\partial S(\hat{\alpha}, \hat{\beta})}{\partial \hat{\beta}} = 0$$

を満たす $\hat{\alpha}, \hat{\beta}$ を求める。
すなわち、 $\hat{\alpha}, \hat{\beta}$ は、

$$\sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0, \quad (1)$$

$$\sum_{i=1}^n X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0, \quad (2)$$

を満たす。
さらに、

$$\sum_{i=1}^n Y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n X_i \quad (3)$$

$$\sum_{i=1}^n X_i Y_i = \hat{\alpha} \sum_{i=1}^n X_i + \hat{\beta} \sum_{i=1}^n X_i^2 \quad (4)$$

(3) 式の辺々を n で割って、

$$\frac{1}{n} \sum_{i=1}^n Y_i = \hat{\alpha} + \hat{\beta} \frac{1}{n} \sum_{i=1}^n X_i$$

すなわち、

$$\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X} \quad (5)$$

を得る。ただし、

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

とする。

さらに、 $\sum_{i=1}^n X_i = n\bar{X}$ と (5) 式を利用して、 $\hat{\alpha}$ を消去すると、

$$\sum_{i=1}^n X_i Y_i = (\bar{Y} - \hat{\beta}\bar{X})n\bar{X} + \hat{\beta} \sum_{i=1}^n X_i^2$$

$\hat{\beta}$ で整理して、

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} \\ &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{XY}}{S_X^2} \end{aligned} \quad (6)$$

が得られ、 $\hat{\alpha}$ は (5) 式から、

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \quad (7)$$

となる。ただし、

$$S_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

とする。

回帰直線は、

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i,$$

として与えられる。 \hat{Y}_i は、 X_i を与えたときの Y_i の予測値と解釈される。

数値例： 以下の数値例を使って、回帰式 $Y_i = \alpha + \beta X_i$ の α, β の推定値 $\hat{\alpha}, \hat{\beta}$ を求める。

i	X_i	Y_i
1	5	4
2	1	1
3	3	1
4	2	3
5	4	4

$\hat{\alpha}, \hat{\beta}$ を求めるための公式は、

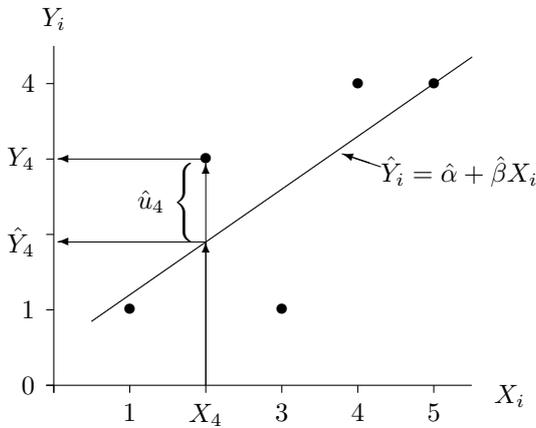
$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X},$$

なので、必要なものは $\bar{X}, \bar{Y}, \sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i Y_i$ である。

i	X_i	Y_i	X_i^2	$X_i Y_i$
1	5	4	25	20
2	1	1	1	1
3	3	1	9	3
4	2	3	4	6
5	4	4	16	16
合計	$\sum X_i$	$\sum Y_i$	$\sum X_i^2$	$\sum X_i Y_i$
	15	13	55	46
平均	\bar{X}	\bar{Y}		
	3	2.6		

表中では、 $\sum_{i=1}^n$ を \sum と省略して表記している。

図 1: $Y_i, X_i, \hat{Y}_i, \hat{u}_i$ の関係



よって,

$$\hat{\beta} = \frac{46 - 5 \times 3 \times 2.6}{55 - 5 \times 3^2} = \frac{7}{10} = 0.7$$

$$\hat{\alpha} = 2.6 - 0.7 \times 3 = 0.5,$$

となる。

注意事項:

1. α, β は真の値で未知である。
2. $\hat{\alpha}, \hat{\beta}$ は α, β の推定値でデータから計算される。

回帰直線は, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ であり, 上の数値例では,

$$\hat{Y}_i = 0.5 + 0.7X_i,$$

となる。 $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_5$ として, 次の表のように計算される。 $Y_i, X_i, \hat{Y}_i, \hat{u}_i$ の関係が図 1 に描かれている。

i	X_i	Y_i	X_i^2	$X_i Y_i$	\hat{Y}_i
1	5	4	25	20	4.0
2	1	1	1	1	1.2
3	3	1	9	3	2.6
4	2	3	4	6	1.9
5	4	4	16	16	3.3
合計	$\sum X_i$	$\sum Y_i$	$\sum X_i^2$	$\sum X_i Y_i$	$\sum \hat{Y}_i$
	15	13	55	46	13
平均	\bar{X}	\bar{Y}			
	3	2.6			

\hat{Y}_i を実績値 Y_i の予測値または理論値と呼ぶ。

$$\hat{u}_i = Y_i - \hat{Y}_i,$$

\hat{u}_i を残差と呼ぶ。 $Y_i, \hat{Y}_i, \hat{u}_i$ の関係, $\hat{Y}_i, X_i, \hat{\alpha}, \hat{\beta}$ の関係は,

$$Y_i = \hat{Y}_i + \hat{u}_i = \hat{\alpha} + \hat{\beta}X_i + \hat{u}_i,$$

の式でまとめられる。

3.3 残差 \hat{u}_i の性質について

$\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta}X_i$ に注意すると, (1) 式, (2) 式から,

$$\sum_{i=1}^n \hat{u}_i = 0, \quad \sum_{i=1}^n X_i \hat{u}_i = 0,$$

を得る。また, $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ から,

$$\sum_{i=1}^n \hat{Y}_i \hat{u}_i = 0,$$

が得られる。なぜなら,

$$\sum_{i=1}^n \hat{Y}_i \hat{u}_i = \sum_{i=1}^n (\hat{\alpha} + \hat{\beta}X_i) \hat{u}_i = \hat{\alpha} \sum_{i=1}^n \hat{u}_i + \hat{\beta} \sum_{i=1}^n X_i \hat{u}_i = 0$$

となるからである。

数値例で確認してみよう。

i	X_i	Y_i	\hat{Y}_i	\hat{u}_i	$X_i \hat{u}_i$	$\hat{Y}_i \hat{u}_i$
1	5	4	4.0	0.0	0.0	0.00
2	1	1	1.2	-0.2	-0.2	-0.24
3	3	1	2.6	-1.6	-4.8	-4.16
4	2	3	1.9	1.1	2.2	2.09
5	4	4	3.3	0.7	2.8	2.31
合計	$\sum X_i$	$\sum Y_i$	$\sum \hat{Y}_i$	$\sum \hat{u}_i$	$\sum X_i \hat{u}_i$	$\sum \hat{Y}_i \hat{u}_i$
	15	13	13	0.0	0.0	0.0
平均	\bar{X}	\bar{Y}				
	3	2.6				

3.4 決定係数 R^2 について

$Y_i, \hat{Y}_i, \hat{u}_i$ の関係は,

$$Y_i = \hat{Y}_i + \hat{u}_i,$$

であった。 \bar{Y} を両辺から引くと、

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + \hat{u}_i,$$

が得られる。さらに、両辺を二乗して、総和すると、

$$\begin{aligned} & \sum_{i=1}^n (Y_i - \bar{Y})^2 \\ &= \sum_{i=1}^n ((\hat{Y}_i - \bar{Y}) + \hat{u}_i)^2 \\ &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + 2 \sum_{i=1}^n (\hat{Y}_i - \bar{Y})\hat{u}_i + \sum_{i=1}^n \hat{u}_i^2 \\ &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{u}_i^2 \end{aligned}$$

となる。二つ目の等式の右辺第二項では、 $\sum_{i=1}^n \hat{Y}_i \hat{u}_i = \bar{Y} \sum_{i=1}^n \hat{u}_i = 0$ が使われている。まとめると、

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{u}_i^2$$

を得る。さらに、両辺を左辺で割ると、

$$1 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} + \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2},$$

が得られる。それぞれの項は、

1. $\sum_{i=1}^n (Y_i - \bar{Y})^2 \rightarrow Y_i$ の全変動
2. $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \rightarrow \hat{Y}_i$ (回帰直線) で説明される部分
3. $\sum_{i=1}^n \hat{u}_i^2 \rightarrow \hat{Y}_i$ (回帰直線) で説明されない部分

となる。

回帰式の当てはまりの良さを示す指標として、決定係数 R^2 が、

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (8)$$

のように定義される。 R^2 は Y_i のうち \hat{Y}_i (または、 X_i) で説明できる比率を意味する。または、

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (9)$$

として書き換えることもできる。

R^2 の取り得る範囲: さらに、 R^2 の取り得る範囲を求める。(8) 式の右辺の分子と分母は共に正なので、 $R^2 \geq 0$ となる。(9) 式の右辺では 1 から第二項の正の値 (分子分母共に正) を差し引いているので、 $R^2 \leq 1$ となることが分かる。すなわち、 R^2 の取り得る範囲は、

$$0 \leq R^2 \leq 1,$$

となる。

$R^2 = 1$ となる場合はすべての i について $\hat{u}_i = 0$ となり、観測されたデータ (X_i, Y_i) は一直線上に並んでいる状態となる。

$R^2 = 0$ となる場合は二通りが考えられる。一つは、 Y_i が X_i に影響されないときで、 $\hat{\beta} = 0$ の状態、すなわち、データが横軸に平行に一直線上に並んでいる状態となる。もう一つは、データが円状に散布していて、どこにも直線が引けない状態である (ちなみに、データが楕円上に散布している場合は、直線が引ける状態である)。

実際のデータを用いた場合は $R^2 = 0$ や $R^2 = 1$ という状況はあり得ない。 R^2 が 1 に近づけば回帰式の当てはまりは良い、 R^2 が 0 に近づけば回帰式の当てはまりは悪いと言える。しかし、「どの値よりも大きくなるべき」といった基準はない。慣習的には、メドとして 0.9 以上が当てはまりが良いと判断する。

データと R^2 との関係は、後述の 3.5 節で、数値例を挙げながら解説する。

R^2 の別の解釈: R^2 のもう一つの解釈をするために、 R^2 の右辺の分子を、

$$\begin{aligned} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y} - \hat{u}_i) \\ &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y}) - \sum_{i=1}^n (\hat{Y}_i - \bar{Y})\hat{u}_i \\ &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y}), \end{aligned}$$

と書き換える。最初の等式では、括弧二乗の一つに $\hat{Y}_i = Y_i - \hat{u}_i$ が用いられている。 R^2 は、

$$\begin{aligned} R^2 &= \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= \frac{\left(\sum_{i=1}^n (\hat{Y}_i - \bar{Y})\right)^2}{\left(\sum_{i=1}^n (Y_i - \bar{Y})\right)\left(\sum_{i=1}^n (\hat{Y}_i - \bar{Y})\right)} \end{aligned}$$

$$= \left(\frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}} \right)^2,$$

と書き換えられる。この式では、 R^2 が Y_i と \hat{Y}_i の相関係数の二乗と解釈されることを意味する。なお、二つ目の等号の右式では、分子と分母に $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$ を掛けていることに注意せよ。

特に、単回帰の場合、 $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ と $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$ を用いて、

$$\begin{aligned} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 &= \hat{\beta}^2 \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \hat{\beta} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}), \end{aligned}$$

を利用すると、

$$\begin{aligned} R^2 &= \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= \frac{\hat{\beta}^2 \sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \\ &= \left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \right)^2 \\ &= \frac{S_{XY}^2}{S_X^2 S_Y^2}, \end{aligned}$$

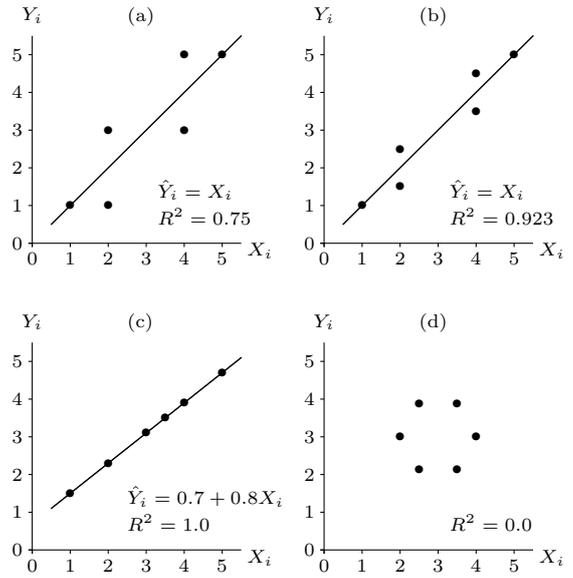
としても書き換えられる。すなわち、単回帰の場合、決定係数は説明変数 X_i と被説明変数 Y_i との相関係数の二乗となる。

数値例： 決定係数の計算には以下の公式を用いる。

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n Y_i^2 - n\bar{Y}^2}$$

計算に必要なものは、 $\sum_{i=1}^n \hat{u}_i^2$, \bar{Y} , $\sum_{i=1}^n Y_i^2$ である。

図 2: 決定係数の比較



i	X_i	Y_i	\hat{Y}_i	\hat{u}_i	\hat{u}_i^2	Y_i^2
1	5	4	4.0	0.0	0.00	16
2	1	1	1.2	-0.2	0.04	1
3	3	1	2.6	-1.6	2.56	1
4	2	3	1.9	1.1	1.21	9
5	4	4	3.3	0.7	0.49	16
合計	$\sum X_i$ 15	$\sum Y_i$ 13	$\sum \hat{Y}_i$ 13	$\sum \hat{u}_i$ 0.0	$\sum \hat{u}_i^2$ 4.3	$\sum Y_i^2$ 43
平均	\bar{X} 3	\bar{Y} 2.6				

$\bar{Y} = 2.6$, $\sum_{i=1}^n \hat{u}_i^2 = 4.3$, $\sum_{i=1}^n Y_i^2 = 43$ なので、

$$R^2 = 1 - \frac{4.3}{43 - 5 \times 2.6^2} = \frac{4.9}{9.2} = 0.5326$$

3.5 決定係数の比較

次の数値例を用いて、決定係数の比較を行おう。 X と Y のプロットしたものが図 2(a)~(d) である。

i	(a)		(b)		(c)		(d)	
	X_i	Y_i	X_i	Y_i	X_i	Y_i	X_i	Y_i
1	1	1	1	1	1	1.5	1	3
2	2	1	2	1.5	2	2.3	2.5	2.134
3	2	3	2	2.5	3	3.1	2.5	3.866
4	4	3	4	3.5	3.5	3.5	3.5	2.134
5	4	5	4	4.5	4	3.9	3.5	3.866
6	5	5	5	5	5	4.7	4	3

(a)と(b)のどちらの場合も、切片・傾きの値は $\hat{\alpha} = 0$, $\hat{\beta} = 1$ として計算されるが、決定係数について、(a)は0.75, (b)は0.923となる(読者はチェックすること)。データのプロットと回帰直線は図2の(a)と(b)に描かれている。 X_i はどちらも同じ数値とした。横軸 X が2, 4のケースについて、(b)が(a)より直線に近くなるように、 Y の値を変えてみた。(b)のデータの方が(a)より直線に近いために、決定係数が0.923と1に近い値となっているのが分かる。(c)はデータが一直線上に並んでいる場合で、決定係数が1となる。決定係数がゼロとなるのは(d)の場合で、 X と Y との関係を表す直線が描けない場合である。(d)の数値例では、 X と Y との関係が円としているが、満遍なく散布している状態と考えてもらえれば良い。

3.6 まとめ

$\hat{\alpha}$, $\hat{\beta}$ を求めるための公式は

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

なので、必要なものは \bar{X} , \bar{Y} , $\sum_{i=1}^n X_i^2$, $\sum_{i=1}^n X_i Y_i$ である。決定係数の計算には以下の公式を用いる。

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n Y_i^2 - n \bar{Y}^2}$$

ただし、 $\hat{u}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$ である。計算に必要なものは、 $\sum_{i=1}^n \hat{u}_i^2$, \bar{Y} , $\sum_{i=1}^n Y_i^2$ である。

4 最小二乗法について：重回帰モデル

k 変数の多重回帰モデルを考える。

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

X_{ji} は j 番目の説明変数の第 i 番目の観測値を表す。 $\beta_1, \beta_2, \dots, \beta_k$ は推定されるべきパラメータである。すべての i について、 $X_{1i} = 1$ とすれば、 β_1 は定数項として表される。 n 組のデータ $(Y_i, X_{1i}, X_{2i}, \dots, X_{ki})$, $i = 1, 2, \dots, n$ を用いて、 $\beta_1, \beta_2, \dots, \beta_k$ を求める。

ある基準の下で、 $\beta_1, \beta_2, \dots, \beta_k$ の解を $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ としよう。データ $\{(X_i, Y_i), i = 1, 2, \dots, n\}$ と直線との関係は、

$$Y_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i = \hat{Y}_i + \hat{u}_i,$$

となる。すなわち、すべての i について、実際のデータ Y_i と直線上の値 $\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$ が一致することはあり得ないので、残差 \hat{u}_i の二乗和を考える。次のような関数 $S(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$ を定義する。

$$S(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k) = \sum_{i=1}^n u_i^2$$

$$= \sum_{i=1}^n (Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

このとき、

$$\min_{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k} S(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$$

となるような $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ を求める。⇒ 最小自乗法最小化のためには、

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)}{\partial \hat{\beta}_1} = 0$$

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)}{\partial \hat{\beta}_2} = 0$$

$$\vdots$$

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)}{\partial \hat{\beta}_k} = 0$$

を満たす $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ となる。

すなわち、 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ は、

$$\sum_{i=1}^n (Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}) X_{1i} = 0,$$

$$\sum_{i=1}^n (Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}) X_{2i} = 0,$$

$$\vdots$$

$$\sum_{i=1}^n (Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}) X_{ki} = 0,$$

を満たす。

さらに、

$$\begin{aligned} \sum_{i=1}^n X_{1i}Y_i &= \hat{\beta}_1 \sum_{i=1}^n X_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^n X_{1i}X_{2i} + \cdots + \hat{\beta}_k \sum_{i=1}^n X_{1i}X_{ki} \\ \sum_{i=1}^n X_{2i}Y_i &= \hat{\beta}_1 \sum_{i=1}^n X_{1i}X_{2i} + \hat{\beta}_2 \sum_{i=1}^n X_{2i}^2 + \cdots + \hat{\beta}_k \sum_{i=1}^n X_{2i}X_{ki} \\ &\vdots \\ \sum_{i=1}^n X_{ki}Y_i &= \hat{\beta}_1 \sum_{i=1}^n X_{1i}X_{ki} + \hat{\beta}_2 \sum_{i=1}^n X_{2i}X_{ki} + \cdots + \hat{\beta}_k \sum_{i=1}^n X_{ki}^2 \end{aligned}$$

の連立方程式を解くことになる。⇒ コンピュータによって計算

4.1 決定係数 R^2 と自由度修正済み決定係数 \bar{R}^2 について

また、決定係数 R^2 についても同様に表される。

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

ただし、 $\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$, $Y_i = \hat{Y}_i + \hat{u}_i$ である。

R^2 は、説明変数を増やすことによって、必ず大きくなる。なぜなら、説明変数が増えることによって、 $\sum_{i=1}^n \hat{u}_i^2$ が必ず減少するからである。

R^2 を基準にすると、被説明変数にとって意味のない変数でも、説明変数が多いほど、よりよいモデルということになる。この点を改善するために、自由度修正済み決定係数 \bar{R}^2 を用いる。

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i^2 / (n - k)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1)}$$

$\sum_{i=1}^n \hat{u}_i^2 / (n - k)$ は u_i の分散 σ^2 の不偏推定量であり、 $\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1)$ は Y_i の分散の不偏推定量である。分散や不偏推定量の意味は、統計学の知識を必要とし、後述する。

R^2 と \bar{R}^2 との関係は、

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k}$$

となる。さらに、

$$\frac{1 - \bar{R}^2}{1 - R^2} = \frac{n - 1}{n - k} \geq 1,$$

という関係から、 $\bar{R}^2 \leq R^2$ という結果を得る。($k = 1$ のときのみ、等号が成り立つ。)

数値例： 今までと同じ数値例で、 \bar{R}^2 を計算する。

i	X_i	Y_i	\hat{Y}_i	\hat{u}_i	\hat{u}_i^2	Y_i^2
1	5	4	4.0	0.0	0.00	16
2	1	1	1.2	-0.2	0.04	1
3	3	1	2.6	-1.6	2.56	1
4	2	3	1.9	1.1	1.21	9
5	4	4	3.3	0.7	0.49	16
合計	$\sum X_i$ 15	$\sum Y_i$ 13	$\sum \hat{Y}_i$ 13	$\sum \hat{u}_i$ 0.0	$\sum \hat{u}_i^2$ 4.3	$\sum Y_i^2$ 43
平均	\bar{X} 3	\bar{Y} 2.6				

$\bar{Y} = 2.6$, $\sum_{i=1}^n \hat{u}_i^2 = 4.3$, $\sum_{i=1}^n Y_i^2 = 43$ なので、

$$\begin{aligned} R^2 &= 1 - \frac{\sum \hat{u}_i^2}{\sum Y_i^2 - n\bar{Y}^2} = 1 - \frac{4.3}{43 - 5 \times 2.6^2} \\ &= 1 - \frac{4.3}{9.2} = 0.5326 \end{aligned}$$

となり、 \bar{R}^2 は、

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{\sum \hat{u}_i^2 / (n - k)}{(\sum Y_i^2 - n\bar{Y}^2) / (n - 1)} \\ &= 1 - \frac{4.3 / (5 - 2)}{9.2 / (5 - 1)} = 0.3768 \end{aligned}$$

となる。

自由度について： 分子について、残差 \hat{u}_i を求めるためには、 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ の k 個の推定値を得なければならない。データ数 n から推定値の数 k を差し引いたものを自由度 (degree of freedom) と呼ぶ。

一方、分母については、 X_{1i} が定数項だとして、 Y_i が定数項を除く $X_{2i}, X_{3i}, \dots, X_{ki}$ に依存しない場合を考える。この場合、 $\beta_2 = \beta_3 = \dots = \beta_k = 0$ とするので、 $\hat{u}_i = Y_i - \hat{\beta}_1$ となる。 \hat{u}_i を得るためには $\hat{\beta}_1$ だけを求めればよい。最小二乗法の考え方に沿って求めれば、 $\hat{\beta}_1 = \bar{Y}$ となる (読者は確認すること)。すなわち、自由度は「データ数 - 推定値の数 = $n - 1$ 」ということになる。

このように、決定係数の第二項目の分子・分母をそれぞれの自由度で割ることによって、自由度修正済み決定係数が得られる。

注意： R^2 や \bar{R}^2 を比較する場合、被説明変数が同じであることが重要である。被説明変数が対数かまたはそのままの値であれば、決定係数・自由度修正済み決定係数の大小比較は意味をなさない。ただし、被説明変数が異なる場合であっても、被説明変数を上昇率とするかそのままの値を用いるかの比較では、決定係数・自由度修正済み決定係数の大小比較はできないが、誤差項 u_i の標準誤差での比較は可能である（標準誤差の小さいモデルを採用する）。⇒関数型の選択

5 ダミー変数

5.1 異常値ダミー

データに異常値が含まれている場合、経済構造がある時期から変化した場合、ダミー変数を使う。

ダミー変数とは、0 と 1 から成る変数のことである。

例えば、データが 20 期間あるとして、9 期目のデータが、回帰直線から離れている場合（異常値の場合）を考える。

$$D_i = \begin{cases} 0, & i \neq 9 \text{ のとき} \\ 1, & i = 9 \text{ のとき} \end{cases}$$

という変数を作り、

$$Y_i = \alpha + \delta D_i + \beta X_i + u_i$$

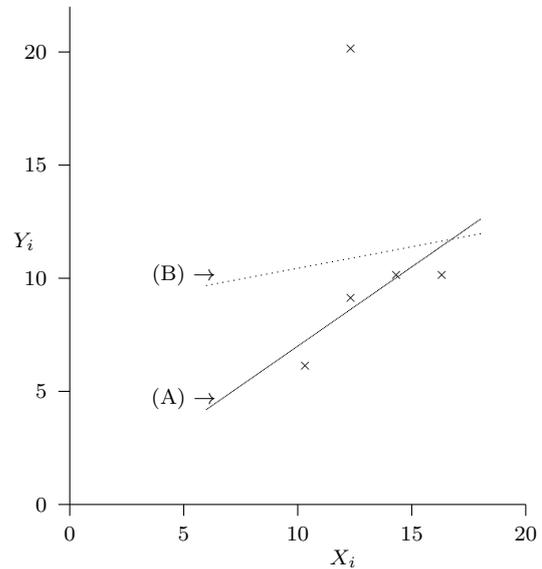
を推定する。 δ の推定値 $\hat{\delta}$ の有意性を調べることによって、異常値かどうかの検定ができる。

数値例： 今までと同様に、以下の数値例をとりあげる。

i	Y_i	X_i	D_i
1	6	10	0
2	9	12	0
3	10	14	0
4	10	16	0
5	20	12	1

第 5 期目が異常値である。

図 3： 異常値



(A) は $i = 1, 2, 3, 4$ のデータを使って、推定した回帰直線である。(B) は $i = 1, 2, 3, 4, 5$ のデータを使って、推定した回帰直線である。(A), (B) の推定結果は以下のとおりである。

$$(A): Y_i = 0.3 + 0.65 X_i,$$

$$R^2 = 0.786, \quad \bar{R}^2 = 0.679$$

$$(B): Y_i = 8.54 + 0.19 X_i,$$

$$R^2 = 0.007, \quad \bar{R}^2 = -0.324$$

ただし、係数の推定値の下の括弧内は t 値を表すものとする。

このように、結果が大幅に変わる。第 5 期は異常値なので、ダミー変数を用いて、

$$Y_i = \alpha + \beta X_i + \gamma D_i + u_i,$$

として推定を行う。 $i = 1, 2, 3, 4$ について、 $D_i = 0$ とし、 $i = 5$ について、 $D_i = 1$ とする変数である。この回帰式の意味は、

$$Y_i = \begin{cases} \alpha + \beta X_i + u_i, & i = 1, 2, 3, 4 \text{ のとき,} \\ (\alpha + \gamma) + \beta X_i + u_i, & i = 5 \text{ のとき,} \end{cases}$$

となる。推定結果は、

$$Y_i = 0.3 + 0.65 X_i + 11.9 D_i,$$

$$R^2 = 0.979, \quad \bar{R} = 0.959$$

となる。この場合、 $\hat{Y}_5 = Y_5$ 、すなわち、 $\hat{u}_5 = 0$ となることに注意。

$$D_{2i} = \begin{cases} 1, & i \text{ が第二四半期のとき} \\ 0, & \text{他} \end{cases}$$

$$D_{3i} = \begin{cases} 1, & i \text{ が第三四半期のとき} \\ 0, & \text{他} \end{cases}$$

という3つのダミー変数を作り、

5.2 構造変化ダミー

次に、9期目以前と以降とで、経済構造が変化している場合を考える。

$$D_i = \begin{cases} 0, & i = 1, 2, \dots, 8 \text{ のとき} \\ 1, & i = 9, 10, \dots, 20 \text{ のとき} \end{cases}$$

という変数を作り、

$$Y_i = \alpha + \delta D_i + \beta X_i + u_i$$

を推定する(定数項だけが変化したと考えた場合)。または、

$$Y_i = \alpha + \delta D_i + \beta X_i + \gamma D_i X_i + u_i$$

を推定する(定数項も係数も変化)。

δ や γ の推定値の有意性を調べることによって、構造変化の検定を行うことができる。

上の例でデータを示すと、

i	Y_i	X_i	D_i	$D_i X_i$
1	Y_1	X_1	0	0
2	Y_2	X_2	0	0
\vdots	\vdots	\vdots	\vdots	\vdots
8	Y_8	X_8	0	0
9	Y_9	X_9	1	X_9
10	Y_{10}	X_{10}	1	X_{10}
\vdots	\vdots	\vdots	\vdots	\vdots
20	Y_{20}	X_{20}	1	X_{20}

となる。

5.3 季節ダミー

季節性のあるデータを扱う場合、

$$D_{1i} = \begin{cases} 1, & i \text{ が第一四半期のとき} \\ 0, & \text{他} \end{cases}$$

5.4 地域差ダミー

関西と関東とで賃金格差があるかどうかを調べたい。

$$w_i = \alpha + \beta D_i + \dots + u_i$$

添え字 i は個人を表すものとする。

$$D_i = \begin{cases} 1, & i \text{ 番目の人が関東に住んでいるとき} \\ 0, & i \text{ 番目の人が関西に住んでいるとき} \end{cases}$$

5.5 男女別ダミー

男女間で賃金格差があるかどうかを調べたい。

$$w_i = \alpha + \beta D_i + \dots + u_i$$

添え字 i は個人を表すものとする。

$$D_i = \begin{cases} 1, & i \text{ 番目の人が女性のとき} \\ 0, & i \text{ 番目の人が男性のとき} \end{cases}$$

6 関数型について

線型：

$$Y_i = \alpha + \beta X_i + u_i,$$

この場合、

$$\beta = \frac{dY_i}{dX_i}$$

なので、 β は、 X_i が一単位上昇(下落)したとき、 Y_i は何単位上昇(下落)するのかを表す。すなわち、 β は限界係数と呼ばれる。

成長率：

$$100 \times \frac{Y_i - Y_{i-1}}{Y_{i-1}} = \alpha + \beta X_i + u_i,$$

として、成長率を被説明変数として用いる場合もある。 $100 \times \frac{Y_i - Y_{i-1}}{Y_{i-1}}$ という変数をあらかじめ作っておき、これをこれまでの Y_i として扱う。

注意：

$$Y_i = \alpha + \beta X_i + u_i \text{ と } 100 \times \frac{Y_i - Y_{i-1}}{Y_{i-1}} = \alpha + \beta X_i + u_i \text{ で}$$

は、得られる決定係数の大きさが全く異なる。単純に、 R^2 や \bar{R}^2 による比較はこの場合出来ない。

$\Rightarrow s^2$ で比較すればよい。

対数線型：

$$\log(Y_i) = \alpha + \beta \log(X_i) + u_i,$$

この場合、

$$\beta = \frac{d \log(Y_i)}{d \log(X_i)} = \frac{\frac{dY_i}{Y_i}}{\frac{dX_i}{X_i}} = \frac{100 \frac{dY_i}{Y_i}}{100 \frac{dX_i}{X_i}}$$

となる。

2つ目の等号では、 $\frac{d \log(Y_i)}{dY_i} = \frac{1}{Y_i}$ が利用される。

3つ目の等号の分子 $100 \frac{dY_i}{Y_i}$ や分母 $100 \frac{dX_i}{X_i}$ は上昇率を表す。

したがって、 β は、 X_i が1%上昇(下落)したとき、 Y_i は何%上昇(下落)するのかを表す。 β は弾力性と呼ばれる。

例：コブ=ダグラス型生産関数：

$$Q_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3}$$

ただし、 Q_i は生産量、 K_i は資本、 L_i は労働である。この場合、対数変換によって、

$$\log(Q_i) = \beta'_1 + \beta_2 \log(K_i) + \beta_3 \log(L_i) + u_i,$$

として、 $\log(Q_i)$ 、 $\log(K_i)$ 、 $\log(L_i)$ のデータをあらかじめ変換しておき、最小二乗法で β'_1 、 β_2 、 β_3 を推定する。また、生産関数には一次同次の制約 $\beta_2 + \beta_3 = 1$ を置く場合が多い。この場合は、

$$\begin{aligned} \log(Q_i) &= \beta'_1 + \beta_2 \log(K_i) + \beta_3 \log(L_i) \\ &= \beta'_1 + \beta_2 \log(K_i) + (1 - \beta_2) \log(L_i) + u_i \\ &= \beta'_1 + \beta_2 (\log(K_i) - \log(L_i)) + \log(L_i) + u_i, \end{aligned}$$

となるので、

$$\log(Q_i) - \log(L_i) = \beta'_1 + \beta_2 (\log(K_i) - \log(L_i)) + u_i,$$

を最小二乗法で推定し、 β'_1 、 β_2 を求めることになる。この場合も同様に、各変数をあらかじめ、 $\log(Q_i) - \log(L_i)$ 、 $\log(K_i) - \log(L_i)$ としてデータを作っておく必要がある。

二次式：

$$Y_i = \alpha + \beta X_i + \gamma X_i^2 + u_i,$$

\Rightarrow 平均費用と生産量との関係等

逆数：

$$Y_i = \alpha + \beta \frac{1}{X_i} + u_i,$$

\Rightarrow 賃金上昇率と失業率との関係(フィリップス曲線)

遅れのある変数： 慣性的効果を考慮に入れたモデル：

$$Y_i = \alpha + \beta X_i + \gamma Y_{i-1} + u_i,$$

ラグ付き内生変数が説明変数に用いられる。

X_i の Y_i への効果は、短期効果、長期効果の2つある。 β は短期効果を表す係数である。長期効果とは、 $Y_i = Y_{i-1}$ となるときの、 X_i から Y_i への影響を示す効果である。

$$Y_i = \alpha + \beta X_i + \gamma Y_i + u_i,$$

として、 Y_i について解くと、

$$Y_i = \frac{\alpha}{1 - \gamma} + \frac{\beta}{1 - \gamma} X_i + \frac{1}{1 - \gamma} u_i,$$

となり、 $\frac{\beta}{1 - \gamma}$ が X_i の Y_i への長期効果を表す係数となる。

問題点：

1. 最小二乗法の仮定の一つに、説明変数は確率変数ではないという仮定がある。ラグ付き内生変数を説明変数に加えることによって、この仮定が満たされなくなる。最小二乗推定量は最小分散線型不偏推定量ではなくなる。
2. Y_i と X_i とは、経済理論的に考えると、相関が高いはず。 Y_i と Y_{i-1} は相関が高い。当然、 Y_{i-1} と X_i も高い相関を示す。
 \Rightarrow 多重共線性の可能性が高い。
3. DW 統計量は意味をなさない。(DW については、後述)

遅れのある変数の解釈 (部分調整モデル) : X_i が与えられたときの Y の最適水準を Y_i^* とする。

$$Y_i^* = \alpha + \beta X_i,$$

現実の水準 Y_i は、最適水準 Y_i^* と前期の水準 Y_{i-1} との差の一定割合と前期の水準 Y_{i-1} との和で与えられるとする。調整関数を考える。

$$Y_i - Y_{i-1} = \lambda(Y_i^* - Y_{i-1}) + u_i,$$

ただし、 u_i は互いに独立で同一な分布の誤差項、 $0 < \lambda < 1$ とする。

よって、

$$Y_i = \lambda\alpha + \lambda\beta X_i + (1 - \lambda)Y_{i-1} + u_i,$$

を得る。

Y_{i-1} と u_i との相関はない。

しかし、 Y_{i-1} が説明変数の一つに入っている (説明変数間が確率変数でないという仮定に反する)。

推定量は不偏推定量ではないが、一致推定量である (証明略)。

7 需要関数の計算と解釈(レポート、締め切り 7月6日 PM17:00 まで 厳守)

レポートの内容は、下記を含めること。

- ・氏名, 学部
- ・何の需要関数を計算したのか?
- ・データの出所
- ・データのグラフ化(推移の説明)
- ・財の数は2つ以上
- ・対数変換で需要関数を推定
- ・各係数の予想される符号(理由も含めて)
- ・得られた結果の解釈(各係数が0以下, 0~1, 1以上)
下級財, 正常財, 上級財, ギフェン財, 必需品, 贅品, 代替財, 補完財などを絡めて説明

7.1 データの入手方法

総務省統計局 <http://www.stat.go.jp/index.htm>

5 家計調査

調査の結果

統計表一覧

家計収支編, 詳細結果表, 二人以上の世帯, 年

2017年(*)

<用途分類> 1世帯当たり年平均1か月間の収入と支出

1-1 都市階級・地方・都道府県庁所在市別 二人以上の世帯 DB

- 「二人以上の世帯のうち勤労者世帯」を選択
- 右横の「更新」をクリック(**)
- その上の「ダウンロード」をクリック
- ファイル形式を「CSV形式」から「XLSX形式」に選択しなおす
- 下の「ダウンロード」をクリック
- 右の「ダウンロード」を再度クリック
- 「ダウンロード」終了後「キャンセル」をクリック
- (**)の画面に戻って、下の方にある「戻る」をクリックすると(*)の画面に戻る

次に、下の方に下記の項目がある

<品目分類> 1世帯当たり年間の支出金額, 購入数量及び平均価格

4-1 全国 二人以上の世帯 DB

→この画面で2か所にDB(一つは「数量」, もう一つは「金額」)が出てくる(***)ので, 下記のように同じ作業を繰り返す

→「二人以上の世帯のうち勤労者世帯」を選択

→右横の「更新」をクリック

→その上の「ダウンロード」をクリック

→ファイル形式を「CSV形式」から「XLSX形式」に選択しなおす

→下の「ダウンロード」をクリック

→右の「ダウンロード」を再度クリック

→「ダウンロード」終了後「キャンセル」をクリック

→(***)の画面に戻って繰り返す

総務省統計局 <http://www.stat.go.jp/index.htm>に戻る。

10 消費者物価指数(CPI)

集計結果

3. 時系列データ

全国(品目別価格指数)

年平均(1970年平均~2017年平均)

中分類指数(1970年~最新年) CSV →「総合」を使う

7.2 例：米の需要関数

年	実収入 (円)	米 (kg)	パン (g)	麺類 (g)	米 (円)	パン (円)	麺類 (円)	消費者 物価指数 (総合)
2000	562754	94.22	41875	36786	37478	29867	19640	99.1
2001	552734	92.81	41124	36967	35899	28878	19196	98.4
2002	539924	87.98	47704	39170	33857	29504	19133	97.5
2003	524810	85.42	49615	39524	33426	29953	18905	97.2
2004	531690	81.57	50373	40520	34496	30262	18865	97.2
2005	524585	82.41	48685	38296	29879	29110	17316	96.9
2006	525719	80.04	48287	37392	28306	28987	16830	97.2
2007	528762	79.15	49775	37674	28115	29840	17129	97.2
2008	534235	80.60	48255	37970	28387	30689	18780	98.6
2009	518226	77.48	49403	38664	27451	31398	19012	97.2
2010	520692	75.41	48859	38871	25987	30591	18625	96.5
2011	510149	74.73	48604	38562	24721	30664	18511	96.3
2012	518506	72.21	48171	37457	26638	30572	18033	96.2
2013	523589	70.74	49183	38069	26225	30708	17823	96.6
2014	519761	65.29	48455	36925	22198	31543	17574	99.2
2015	525669	64.40	49205	36934	20424	32927	18389	100.0
2016	526973	64.50	48468	36505	21412	32772	18198	99.9
2017	533820	63.30	47697	35384	22095	32058	17679	100.4

- 実収入 (円) = 二人以上の世帯のうち勤労者世帯の1世帯当たり年平均1か月間の実収入 (Y)

→ 名目データ

- 米 (kg) = 1世帯当たり年間の購入数量 (Rkg)
 - パン (g) = 1世帯当たり年間の購入数量 (Bg)
 - 麺類 (g) = 1世帯当たり年間の購入数量 (Ng)
 - 米 (円) = 1世帯当たり年間の支出金額 (R)
- 名目データ

- パン (円) = 1世帯当たり年間の支出金額 (B)
- 名目データ

- 麺類 (円) = 1世帯当たり年間の支出金額 (N)
- 名目データ

- 消費者物価指数 (総合) = 2015年基準 (CPI)

(*) 名目データ = 観測された元のデータ

実質データ = ある時点の貨幣価値で測ったデータ

手順

1. 実収入 (円) を実質データに変換する

$$\rightarrow \frac{Y}{CPI/100} \rightarrow \text{グラフ化}$$

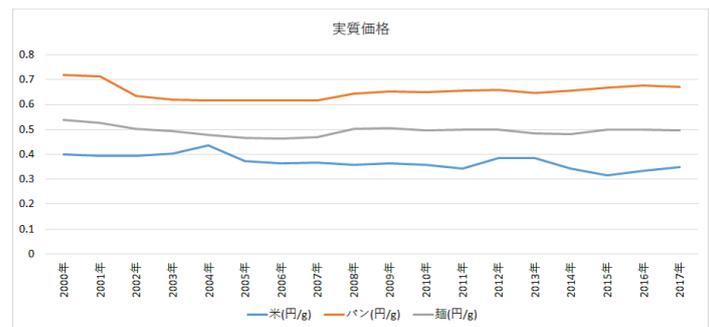
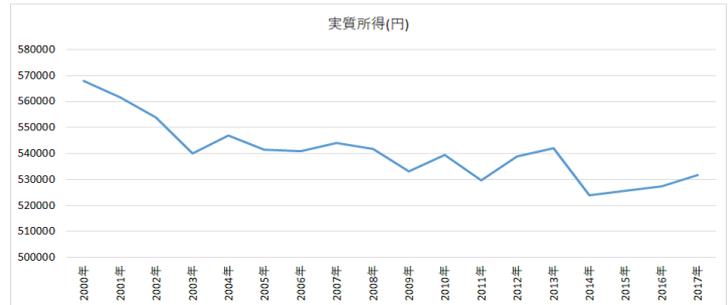
2. それぞれの実質価格データを作る (円/g)

$$\rightarrow \frac{R}{1000 \times Rkg \times CPI/100}$$

$$\frac{B}{Bg \times CPI/100}$$

$$\frac{N}{Ng \times CPI/100}$$

→ グラフ化



3. まずは、これら4つのデータを使って、回帰分析を行う。

回帰統計						
重相関 R	0.927794					
重決定 R2	0.860802					
補正 R2	0.817972					
標準誤差	4037.74					
観測数	18					
分散分析表						
	自由度	変動	分散	F 値	有意 F	
回帰	4	1.31E+09	3.28E+08	20.0981	1.79E-05	
残差	13	2.12E+08	16303343			
合計	17	1.52E+09				
	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	-299861	55692.04	-5.38427	0.000124	-420176	-179546
X 値 1	0.740014	0.152296	4.859039	0.000312	0.410997	1.06903
X 値 2	-30027.47	61308.57	-0.48978	0.632454	-162477	102421.7
X 値 3	-183339.9	76367.99	-2.40074	0.03204	-348323	-18356.9
X 値 4	218155.2	116815.4	1.867521	0.084535	-34209.1	470519.4

$$Rg = -299861.0 + 0.7400Y - 30027.5R - 183340.0B + 218155.2N$$

$$R^2 = 0.8608, \text{ 推定期間 } 2000 \sim 2017 \text{ 年 } (n = 18)$$

上式の Rg, Y, R, B, N はそれぞれ $1000 \times Rkg, \frac{Y}{CPI/100}, \frac{1000 \times Rkg \times CPI/100}{R}, \frac{B}{Bg \times CPI/100}, \frac{N}{Ng \times CPI/100}$ を表す

さらに、それぞれ4つのデータを対数変換して、回帰分析を行う。

回帰統計						
重相関 R	0.925672					
重決定 R2	0.856868					
補正 R2	0.812827					
標準誤差	0.053287					
観測数	18					
分散分析表						
	自由度	変動	分散	F 値	有意 F	
回帰	4	0.220985	0.055246	19.4563	2.14E-05	
残差	13	0.036913	0.002839			
合計	17	0.257898				
	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	-57.17594	15.01682	-3.80746	0.002176	-89.6178	-24.7341
X 値 1	5.194201	1.111271	4.674108	0.000435	2.793446	7.594957
X 値 2	-0.154827	0.310744	-0.49825	0.626633	-0.82615	0.516495
X 値 3	-1.721929	0.664757	-2.59032	0.022414	-3.15805	-0.28581
X 値 4	1.462989	0.755681	1.935988	0.074918	-0.16956	3.095538

$$\ln(Rg) = -57.2 + 5.19 \ln(Y) - 0.155 \ln(R) - 1.722 \ln(B) + 1.463 \ln(N)$$

$$R^2 = 0.8569, \text{ 推定期間 } 2000 \sim 2017 \text{ 年 } (n = 18)$$

● 対数変換の結果：

- 所得↑ ⇒ 米需要量↑
(係数が5.19と1より大きいので、米は景気に左右されやすい財)
- 米価格↑ ⇒ 米需要量↓
(需要法則は成り立っている。価格変動に対して、需要量の変化は小さい。→ 必需品)
- パン価格↑ ⇒ 米需要量↓ (米とパンは補完財)
- 麺類価格↑ ⇒ 米需要量↑ (米と麺類は代替財)

● Word ファイルを添付して、下記のアドレスにメールでレポート提出

tanizaki@econ.osaka-u.ac.jp

Subject: には C クラスの「C」と「学籍番号」を書くこと。
メールの本文は何も書かなくてよい。